Report of the working group on methods for averaging surveys: *Updated through 2013*

Background

For the 2012 assessment cycle, the Groundfish Plan Teams appointed a number of working groups. General terms of reference for the working groups were as follow:

> "Working groups are tasked with developing/collating and analyzing alternative policies/approaches for their respective topic. Analyses can be either quantitative or qualitative (i.e., listing likely pros and cons). Ideally, the working group reports will be substantial enough that the Teams can use them to make informed policy recommendations, which will then be forwarded to the SSC for comment."

In 2012 the SSC reviewed the first version of the draft and had the following comments:

> *"The SSC encourages authors or the GPT to document the Kalman filter (KF) and random effects (RE) models that were proposed for use in assessments. The inclusion of equations describing the models can help reviewers identify the structure of errors in the observation and state equations. Identification of over-parameterization in the KF approach is very difficult, so the authors should check whether they have sufficient replicates and data for their proposed model.*
>
> *"The Discussion section of the report could be strengthened to include a more general discussion of the advantages and disadvantages of the alternative weighting methods, so that the recommendations do not appear to depend so strongly on a single simulation study. For example, it is worth noting that in general, bias will increase with increasing weight given to past observations when there is a trend in the data, and that this is a particularly undesirable property of the equal-weighting methods. Precision, on the other hand, will generally improve as more data are included. The KF essentially balances bias and precision, leading to estimates that are both more precise than using a single survey, but generally have relatively little bias compared to more naive weighting methods. In addition, the KF approach can model process errors, measurement errors and random effects into one likelihood that is free of high dimensional integrals. The RE models usually help the authors to understand the correlation of two random effects and the prediction ability of RE models is the same as the fixed effects models.*

With these comments in mind, this report describes the working group activities to date. The work was conducted by Jim Ianelli, Paul Spencer, Grant Thompson, and Jon Heifetz. Specific topics assigned to the group included the following:

1. Methods for using survey time series to produce a "reliable" estimate of biomass for stocks/complexes managed under Tier 5, including an inventory of methods presently in use.
2. Methods for using survey time series to apportion ABC among areas.
3. Methods for filling in unsurveyed areas during years when survey funding was unavailable. For example, this applies to the groundfish bottom trawl survey in the GOA, and some periods for the sablefish longline survey (e.g., where the eastern Aleutian area is surveyed in alternate years with EBS slope areas).

The working group focused on topics #1 and #2. Further work on topic #2 is needed and work on topic #3 was initiated based on an extension of topic #1. In 2013 the simulations were updated and most notably some refinements and new methods were introduced as described below.

Simulation modeling

Two simulation models were developed in order to generate data sets for testing various methods for obtaining reliable biomass estimates and subarea proportions from survey time series. First, a single-area model was developed to evaluate estimation of biomass. Operational models were developed for "Pacific ocean perch" (POP) and "walleye pollock" life-history patterns. The group considered the following variables in conducting the simulations: coefficient of variation (CV) of survey biomass estimates, survey frequency/data availability, stock longevity/productivity, trend in fishing mortality/biomass, and recruitment variability. The single-area model simulations used the following parameter settings:

Survey CV:
This was approximated by the σ parameter of a lognormal distribution, and values of 0.15 and 0.35 were used.

Natural mortality (M):
Set to 0.06 and 0.30 for the POP and walleye pollock life history types, respectively.

Recruitment variability (σ_R):
Evaluated at 0.8 and 0.4.

Survey frequency:
We considered annual, biannual, and triannual survey schedules.

Trend in fishing rate/biomass:
Cyclic trends in biomass were produced by varying the fishing rate over time. Three patterns were evaluated:

1) The fishing rate was specified to increase over the first half of the time series then decrease: F_{spr} changes from $F_{100\%}$ to $F_{20\%}$, then from $F_{20\%}$ to $F_{100\%}$;
2) The fishing rate was specified decrease over the first half of the time series, then increase: F_{spr} changes from $F_{20\%}$ to $F_{100\%}$, then from $F_{100\%}$ to $F_{20\%}$; and
3) The fishing rate was held constant: F_{spr} set to $F_{50\%}$ for the entire period.

For these three fishing mortality patterns the final portion of the simulation period (which affects the management control rule) biomass trends were increasing, decreasing, and flat, respectively.

The variables above result in 36 permutations for each life history type. For each permutation, 100 simulations of 54 years were conducted (this ensured that the end year had a survey for each of the three survey frequencies).

Actual trawl surveys produce estimates of both the survey biomass and the variance of the estimated survey biomass. In the simulations, the survey biomass estimates *and their variances* were treated as random variables based on the samples observed in a particular survey. Simulated estimates of the variance in survey biomass were assumed to have a lognormal distribution with the mean set to the "true variance" and standard deviation of 0.15.

Simulations with movement

The work on simulation models with movement is unchanged since the 2012 draft. For analysis of estimation of area proportions, the single-area model was expanded to a three-area model, with the areas organized in a linear pattern (as along a coastline or island chain). At each time step, the total number of recruits was obtained from a Beverton-Holt recruitment function applied to the total level of spawning stock biomass. The total predicted recruits were distributed with 40% to the central area and 30% to each of the other two areas. Recruitment variability was then added to the predicted recruits in each subarea, and this variability incorporated global variability (identical for all subareas) and local variability (separate for each subarea). The variability was modeled with lognormal distributions, with the global values of σ evaluated at 0.8 and 0.4, and the local values of σ were set to one half the global value.

Adults were allowed to move between areas. The proportion of adults moving from a subarea was modeled as function of age with a logistic function. Two levels for maximum proportion of fish (by age) moving were used (0.1 and 0.3). The age at which the movement rate reached 50% of maximum was set to age 3 for pollock and age 8 for POP; these ages roughly correspond to the age at 50% maturity.

In the multiple-area model, estimates of survey biomass are modeled for each subarea. Because of this finer scale of resolution, the subarea survey CVs were increased from those used in the single area model. Values of 0.25 and 0.6 were evaluated.

Evaluation of the two values of maximum proportion of adult movement increased the number of permutations for each life-history type to 72. As with the single-area model, 100 simulations of 54 years were conducted for each permutation.

Finally, the trend in fishing biomass for the simulation with movement used monotonic patterns in F over the simulated time period: 1) an increasing biomass trend (F_{spr} changes from $F_{20\%}$ to $F_{100\%}$ over the simulation); 2) a decreasing biomass trend (F_{spr} changes from $F_{100\%}$ to $F_{20\%}$ over the simulation); and 3) a constant trend (F_{spr} held at $F_{50\%}$ over the simulation).

Estimation methods

The first task completed was an inventory of methods currently used for averaging survey estimates for stock assessment and ABC considerations for the BSAI and GOA (Table 1). The totals shown in Table 1 are broken down by region and stock/complex in Table 2. These tables illustrate the variety of methods presently in use and also provided guidance on a range of methods to evaluate against the simulated data:

1. The most recent survey estimate
2. Simple recent N survey average (e.g., N= 3, 4)
3. Exponential weighting (EW) with weights increasing for more recent surveys
4. Simple random effects (RE) model, with trend assumed to be zero (i.e., random walk). Note that the simple random effects model is the same as a Kalman filter if the transition and observation equations are linear and the process and observation errors are uncorrelated and normally distributed.

In the BSAI, methods for obtaining biomass for Tier 5 stocks include all of the methods listed above, with a simple average and the most recent survey biomass estimate being the most common. For GOA Tier 5 stocks, a simple average was the most common method, being used in 7 of 11 cases. In the BSAI, area apportionments are used in six cases, of which 4 use a weighted average. In contrast, area proportions were used in 19 cases in the GOA, with the most common methods being weighted average (7 cases) and most recent survey (7 cases).

The following methods applied to the simulated survey data, and were motivated by methods currently used.

1) Exponential smoothing
2) Simple Kalman filter (random walk with observation error)
3) Simple random effects model (random walk with observation error)

4) A generalized ARIMA modeling method (Stockhausen and Fogarty 2007).

The exponential smoothing model can be expressed as

$$\hat{x}_t = \hat{y}_t(1) = (\alpha)y_t + (1-\alpha)\left[\alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \dots\right]$$

with α is an exponential smoothing parameter, *y* is the survey biomass, and *x* is the underlying "true" survey biomass (i.e., without observation error). The exponential smoothing parameter is a function of the ratio of observation over process error variances. For random walk models with constant variances exponential smoothing can be shown as optimal for forecasting (Pennington 1986). In this case, exponential smoothing corresponds to applying a Kalman filter or a simple ARIMA time series model (see below for a description of these methods), and either of these methods could be used to estimate the exponential smoothing parameter. Several of the currently used methods (i.e., taking the most recent survey estimate, and taking either weighted or unweighted averages survey estimates) correspond to particular values of the smoothing parameter, with the important distinction that exponential smoothing would be applied to the entire time series rather than the most recent points. A low exponential weight implies you give a low weight in the current biomass and relatively more weight to historical surveys (but with the weights still tapering for older surveys).

The RE approach applied here partitions the variability due to underlying processes (i.e., changes in population "state" from one year to the next) from the observation errors due to sampling. The RE model will estimate high values for process errors when the population appears to fluctuate broadly and observations are highly precise. In situations where observation errors are large, the ability to detect population fluctuations will decrease and the overall uncertainty may remain high.

A description of the transition and observation equations used in the RE model follows:

Notation: x_0 is the true (log) survey biomass at time 0, and x_t is the true (log) survey biomass at time *t*, σ_p is the process error standard deviation, y_t is the observed (log) survey biomass at time *t*, and $\sigma_{o,t}$ is the standard error of the survey observation at time *t*. Then, the transition and observation equations can be written as:

$$x_t = x_{t-1} + \varepsilon_{p,t-1} \quad \text{for } t = 1, 2, \dots, n, \text{ where } \varepsilon_{p,t} \sim N\left(0, \sigma_p^2\right)$$

$$y_t = x_t + \varepsilon_{o,t} \quad \text{for } t = 1, 2, \dots, n, \text{ where } \varepsilon_{o,t} \sim N\left(0, \sigma_{o,t}^2\right)$$

The RE model gives the marginal likelihood after all of the x_t have been integrated out. The two parameters estimated by maximizing the marginal likelihood are x_0 and σ_p. The distributions of the x_t are given in closed form after running the Kalman "smoother," conditional on the point estimates of x_0 and σ_p. The Kalman filter requires linear dynamics and Gaussian error distributions; lognormal error distributions can be used by using the natural log of survey biomass as the input data.

As noted in the 2012 draft, an approximation to the Kalman Filter can be written as a random effects (RE) model where the process errors (step changes) from one year to the next are the random effects to be integrated over and the process error variance is a free parameter. The observations can be irregularly spaced. The box below shows the contents of a typical data file:

```
# Aleutian Islands Kamchatka flounder
# Year range
1991 2013
#Number of observations
8
#Years of observations
 1991  1994  1997  2000  2002  2004  2006  2010
#Biomass estimates
16255 49156 37664 28535 49035 39219 45369 53962
#Std Errors of biomass estimates
4458  18522  9588  6601 13634  9219 11058 20567
```

The KF and RE-normal models are exactly the same in principle. The only difference, in practice, is that the KF model integrates the joint likelihood (to obtain the marginal likelihood) exactly, by exploiting certain convenient properties of the linear-normal equations, whereas the RE model integrates the joint likelihood using a very accurate approximation, which allows it to move beyond the linear-normal case if desired. In Skaug and Fournier (2006; referenced in Fournier et al. 2012), the last paragraph on page 708 discusses a logistic growth model with linear-normal random effects and linear-normal error, and notes, “For this model the Laplace approximation is exact (y_{ij} is normally distributed).” The random effects model has the additional advantage of flexibility whereas the Kalman filter assumes normal error distributions and linear dynamics.

The ARIMA (auto-regressive integrated moving average) model applied here can be expressed as

$$y_t = \alpha_1 y_{t-1} + \ldots \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-q} \ldots \beta_q \varepsilon_{t-q}$$

The structure of the ARIMA model is denoted “(p,d,q),” where p is the order of the autoregressive part, d is the number of times the data are differenced (for stationarity) and q refers to the order of the moving average process. For example, the random walk plus uncorrelated noise (RWPUN) model would be denoted ARIMA(0,1,1).

The advantage of this approach is relaxation of the assumption that the underlying process is a random walk. Following Stockhausen and Fogarty (2007), the smoothing procedure was based on fitting generalized ARIMA models:

1) Fit a series of candidate ARIMA models to survey data.
2) Use model selection criteria to identify the best p,d,q ARIMA model.
3) Estimate the power spectrum for the ARIMA process, which gives an estimate of the upper bound on the observation error variance (K*).
4) From ARIMA parameters and K*, estimate smoothing weights to be used in a symmetric moving average.

Conditions for applying generalized ARIMA smoothing:

1) A time series long enough to get reliable parameter estimates (Stockhausen and Fogarty (2007) suggested 40 years)
2) Estimated $Q \geq$ (P+D)
3) Signal other than pure white noise
4) Other (stationarity of autoregressive parameters, invertability, variance reduction)

Our simulations occurred over 54 years and satisfy condition 1. Condition 2 results from fitting the generalized ARIMA model for the underlying “true” survey biomass being contained within ARIMA model for the observed time series. Condition 3 means that the smoothing must vary according to autocorrelation and moving average terms as distinguished from “white noise.” Failing this means that process and observation error variances are indistinguishable. The generalized ARIMA approach of Stockhausen and Fogarty (2007) requires that these conditions be met and thus only worked on a subset of simulations.

Statistics selected to evaluate the performance of the various methods include the mean relative error of biomass (relative error is defined here as estimate/true-1) and variability in these relative errors.

Results

Estimation of survey biomass

For the pollock-like stock, the mean relative error (averaged across runs for each scenario) for a subset of the different simulation scenarios and estimation methods indicate that the RE model performs well in most cases but the absolute value of the bias for all methods was less than 5% in 8 out of 12 scenarios (Table 3). In a relative sense, the RE model performed worse for the case where σ_R and σ_{survey} was high and the biomass was increasing (fishing mortality was declining). Note that the direction of the bias was positive indicating that, as with the other methods, the RE model tended to “flatten out” the biomass trend. The variability of the mean relative error for the pollock model was also relatively low for the RE model with the lowest variability in 8 out of 12 cases (Table 4).

For the rockfish-like simulations were slightly less favorable for the RE model but the differences between estimation methods were similar when the mean absolute relative error was greater than 5% (which occurred in 4 out of 12 cases for the subset shown in Table 5). In 6 of the 12 simulation scenarios presented the RE model had the lowest variability (Table 6).

For the generalized ARIMA model, the “good” cases were mostly random walks. Cases which failed the conditions were from several reasons most commonly that they were categorized as being “white noise only.” They also failed in some cases because the model structure did not meet the requirement of the moving average order (Q) being at least as large as the sum of the autoregressive order (P) plus the degree of differencing (D). In other cases the variance of the smoothed estimates was greater than the original data (i.e., smoothing was unnecessary; Fig. 1). Runs producing white noise were more prevalent with the flat trajectory.

When comparing the generalized ARIMA method with the other ARIMA (0,1,1) methods for rockfish-like scenarios (the subset being with surveys in every year since there was little difference in relative terms to the other approaches) the two methods were nearly equivalent. The generalized ARIMA method shows a bit more variability, in part because it did "less" smoothing due to low estimates of the observation error variance (Fig. 2). The generalized ARIMA model performs about as well as the EW and RE models. The random walk model (i.e., ARIMA (0,1,1)) described many of our datasets. For these cases, the three smoothing methods perform similarly. For a number of datasets, the generalized ARIMA smoothing failed to provide estimates. If the best ARIMA model is other than a (0,1,1), then the generalized ARIMA smoothing could reduce the bias but may increase the variance of the estimated error.

Choosing the EW parameter *a priori* imposes a large probability of the resulting estimator being a poor one. Estimating the optimal value of the exponential weighting parameter, and then applying it *a posteriori*, will likely result in a reasonably performing estimator (especially if the process is truly a random walk). As such, estimators with fixed weights will perform poorly (except when the fixed weight coincidentally matches the estimated weight).

Discussion

Webster (2011) conducted an evaluation (without simulations) using a Kalman filter approach with alternative models including the trendless random walk used here along with 3 other forms that allowed for underlying trends to be estimated. He concluded that the trendless model performed adequately and, based on evaluations of halibut longline survey data from a variety of different areas, that a general historical weighting scheme was a suitable approximation to the results from the Kalman filter (with the 3-most recent surveys being weighted 70:25:5, with 70 being the most recent). The RE method applied here could be used to develop similar "rules of thumb" but this may likely vary by species.

Under certain conditions, EW corresponds to a Kalman Filter, which can also be replicated with the RE model. The optimal vale of the EW parameter can be estimated (for example, with an ARIMA (0,1,1) model), but this process is at least as complicated as applying the RE model.

One distinguishing feature between the RE and ARIMA models is that the RE model equates the observation error variance in each survey year with the sampling variance estimated by the RACE division for that year's survey, while the ARIMA model estimates a single (i.e., time-invariant) observation error variance internally. It is unclear how best to interpret annual sampling variance estimates. On one hand, it could be argued that, if changes in estimated survey variance reflected changes in catchability or availability, then perhaps an average (i.e., constant) variance for survey observation errors would be inappropriate. On the other hand, given that the state variable in the RE model is true survey biomass, changes in catchability or availability are part of process error, and it could be argued that accounting for such changes in the observation error variance would be better.

Contrasting the generalized ARIMA approach with the other two approaches, if the best fitting ARIMA model is a random walk, then the three methods appear to give similar results. If the best fitting ARIMA model is something other than a random walk, the generalized ARIMA smoothing procedure appears to favor "less" smoothing and produces estimates closer to the observed values. This could reduce the bias, but at the expense of increased variance.

The generalized ARIMA smoothing fails to apply in all cases because it depends on having trends in the time series. Thus, from a pragmatic standpoint, methods which continue to function even when the data appear to have only "white noise" might be the best choice. Implementing the generalized ARIMA modeling and smoothing procedure also appears to present additional complexity and may be more challenging to make available for routine use.

Reflecting on the tasks of 1) obtaining a "reliable" estimate of biomass for Tier 5 stocks, 2) evaluating methods for using survey time series to apportion ABC among areas, and 3) evaluating methods for covering unsurveyed areas we note that for Tier 5 stocks, the RE model could be used to address these topics simultaneously. The RE model could be applied to each area separately which would enable calculation of apportionments and filling in for unsurveyed areas. The overall ABC could then be based on the sums of the individual areas.

To examine qualitatively how the RE model would apply results are provided for some selected Aleutian Islands stocks (Fig. 3). Note the interplay between the magnitude of the observation errors and that predicted by the model and how this can provide a way to naturally weight observations going forward.

This model can also be applied to situations where there are missing regions in some years as is the case with the GOA. Figures 4 and 5 show two stocks where the model is fit to each region independently, noting that the values for 2001 are missing due to lack of funding to complete a GOA-wide survey. Given that the performance of the KF method tested well for individual time series, it may follow that applying it to regional time series to "fill in" missing years of data may be appropriate (and it should be feasible to compute uncertainties for application in P^*-based ACLs).

Future work

The main purpose of obtaining a good estimate of current survey biomass is to use this estimate in the Tier 5 control rules. Thus, another issue that should be investigated further is the uncertainty, or possible error, associated with these control rules. In the case of OFL, the Tier 5 control rule, as usually applied, gives: $\mathrm{OFL}_{y+1} = M \times B_y$. In order for this formula to be a good approximation of the one-year-ahead catch at F=F_{MSY}, two conditions need to be met (or at least the degrees to which they are not met need to be offsetting):

- The best estimate of M is good approximation for F_{MSY}.
- The best estimate of survey biomass in year y is a good approximation for mean exploitable biomass during year y+1 if the stock were exploited at F=M throughout year y+1.

Focusing initially on the first condition, some of the literature suggests that the ratio of M to F_{MSY} can easily diverge significantly from unity (e.g., Thompson, 1992, *Fish. Bull.* 90:552-560; Thompson, 1993, *Fish. Bull.* 91:718-731 (errata *Fish. Bull.* 92(3):iii)). Recent analysis (unpublished) has tended to confirm this conclusion. Even when central tendencies of population dynamic parameters are constrained to follow the M=F_{MSY} rule of thumb, a modest amount of variability around those central tendencies can cause the rule to err significantly. For example, when the “simple” model of Beverton and Holt (1957) is combined with a Beverton-Holt stock-recruitment relationship (with a 60% coefficient of variation in recruitment), coefficients of variation on the order of 20% in the model parameters give a 95% confidence interval around the M/F_{MSY} ratio extending from about 0.5 to 2.0.

One difficulty in meeting the second condition is that the assumption of equivalence between survey biomass and exploitable biomass can break down easily, as was noted during this year’s CIE review of the Tier 5 groundfish assessments. For example, fishery selectivity might not equal survey selectivity, and survey catchability might not equal unity.

One recent development that may help to address both of the above problems, but which needs further testing, is the “survey/exploitation vector autoregressive” (SEVAR) model, which debuted this summer at the World Conference on Stock Assessment Methods. The only data requirements for the SEVAR model are time series of survey biomass (either relative or absolute) and total catch, with standard errors for both. The ratio of total catch to survey biomass is used as the measure of exploitation. After some rescaling, the state variables in the SEVAR model consist of true survey biomass and true exploitation rate, which are stacked in a vector. The transition equation is linear and autoregressive with normal error, and the observation equation is also linear and normal. For a model with p time lags, 4(p+1) parameters need to be estimated. The choice of p can be based on Schwarz’s information criterion or similar statistic. After rearranging some terms, the SEVAR model can be cast as a Kalman filter, meaning that the state variables are integrated out automatically, thereby improving accuracy of parameter estimates. Two of the advantages of the SEVAR model are: 1) because survey biomass and exploitation rate covary in the model, the MSY exploitation rate can be estimated directly rather than assumed; and 2) because exploitation is defined with survey biomass as the denominator, projecting the harvest corresponding to the MSY exploitation rate for some future year involves projection of survey biomass only (not exploitable biomass), meaning that neither selectivity nor catchability appear as parameters in the model, and so do not need to be estimated.

Only some of the methods used hear apply the “within survey variance” (year-specific estimates). It is unclear how best to interpret annual variance estimates. One suggestion is that if changes in estimated survey variance reflected changes in the catchability or availability, then perhaps an average constant CV for survey observation errors would be appropriate (as opposed to if the variability was just due to sampling).

Area apportionment continues to be an important topic and will be the focus of extending this work in the coming year.

Recommendations

We recommend that the RE model be applied to obtain the "reliable biomass" estimate required for Tier 5 stocks and applied for the 2013 assessments. Software is presently available from the working group. The following lists some advantages and disadvantages of this approach with respect to this model:

Pros: 1) It is simple to apply. 2) It has a strong theoretical basis and requires minimal conditions for estimates to be available (e.g., if the pattern of data are indistinguishable from a "white-noise" only situation); 3) It performed well in the simulations; 4) Application to actual data showed favorable characteristics for most species; 5) It likely will also be appropriate for computing apportionments. 6) It provides estimates of biomass variances which combine the process errors and observation errors (which may be useful for application to risk-averse ACL specifications). 7) It can provide insight on the loss of information as surveys become more or less infrequent.

Cons: 1) other methods had better performance for some cases (e.g., for stocks that were increasing). 2) This method is more complex than applying simple mean (or weighted mean) values. 3) May perform poorly for stocks or stock complexes that are rarely observed (i.e., which may be absent from the surveys in some years). 4) as with the other methods evaluated here, estimates tend to be biased low for situations where the stock is increasing.

References

Fournier, D.A., H.J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M.N. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27:233-249.

Skaug, H. J., & Fournier, D. (2006). Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models, Comput. Stat. Data Anal. 56 (2006), pp. 699–709.

Stockhausen W. and W. Fogarty. 2007 Removing observational noise from time series data using ARIMA models. Fish. Bull. 107:88–101 (2007).

Thompson, G. G. 1993. Management advice from a simple dynamic pool model. *Fish. Bull.* 91:718-731

Thompson, G. G. 1992. A Bayesian approach to management advice when stock-recruitment Parameters are uncertain. *Fish. Bull.* 90:552-56.

Webster, R. 2010. Weighted averaging of recent survey indices. http://www.iphc.int/papers/SurveyWeighting_web.pdf

Tables

Table 1. Summary of methods used for different stocks by tiers for the BSAI and GOA.

Bering Sea Aleutian Islands

Tier:	1	2	3	4	5	6
Number of stocks	3	0	12	0	7	3
Biomass estimation method						
NA	3		11			3
Average			1		3	
Weighted average					1	
Kalman filter					1	
Most recent					2	
Proportion estimation method						
NA	3		7		6	3
Average			1			
Weighted average			4		1	

Gulf of Alaska

Tier	1	2	3	4	5	6
Number of stocks	0	0	9	2	11	5
Biomass estimation method						
NA			9			5
Average				1	7	
Most recent				1	3	
Mature biomass from model					1	
Proportion estimation method						
NA				1	3	4
Average			2		2	
Weighted average			4	1	2	
Most recent			3		4	
Proportion of historical catch						1

Table 2. Inventory of methods used for different stocks which involved some form of survey averaging (unless otherwise specified).

Area	**Stock**	**Tier**	**Biomass**	**Area proportions**
BSAI	Pacific cod	3	Scaled to BSAI using Kalman filter	NA
AK	Sablefish	3	NA	5 year weighted average of survey and fishery indices
BSAI	Greenland Turbot	3	NA	Most recent three
BSAI	POP	3	NA	4-6-9 weighting by subarea
BSAI	Rougheye/BS	3	NA	4-6-9 weighting by subarea
BSAI	Alaska skate	3	NA	NA
BSAI	Atka mackerel	3	NA	8-12-18-27 weighting
BSAI	Kamchatka	5	7-year average	NA
Bogo	Pollock	5	Most recent	NA
BSAI	Other flatfish	5	Most recent	NA
BSAI	Shortraker rockfish	5	Kalman filter	NA
BSAI	other rockfish	5	4-6-9 weighting	4-6-9 weighting
BSAI	other skates	5	Most recent three	NA
BSAI	sculpins	5	Most recent three	NA
GOA	pollock	3	NA	4 most recent average
GOA	Pacific cod	3	NA	3 most recent average
GOA	Arrowtooth	3	NA	Most recent
GOA	flathead sole	3	NA	Most recent
GOA	northern rockfish	3	NA	4-6-9 weighting
GOA	Pel. shelf rockfish (dusky)	3	NA	4-6-9 weighting
GOA	POP	3	NA	4-6-9 weighting
GOA	RE/BS rockfish	3	NA	4-6-9 weighting
GOA	Shallow flats N, S rock sole	3	NA	Most recent
GOA	Demersal shelf	4	Most recent	NA
GOA	Other rockfish - sharpchin	4	Most recent three	4-6-9 weighting
GOA	Big skate	5	Most recent three	Most recent three
GOA	deep flats Dover sole	5	most recent	Most recent (dover)
GOA	longnose skate	5	Most recent three	Most recent three
GOA	Oher rockfish - other	5	Most recent three	4-6-9 weighting
GOA	Other skates	5	Most recent three	NA
GOA	rex sole	5	Mature biomass from model	Most recent
GOA	Sculpins	5	Most recent four	NA
GOA	shallow flats - others	5	Most recent	Most recent
GOA	Sharks - spiny dogfish	5	Most recent three	NA
GOA	shortraker rockfish	5	Most recent three	4-6-9 weighting
GOA	Thornyhead	5	Most recent	Most recent
GOA	Atka mackerel	6	NA	NA
GOA	deep flats others	6	NA	Proportion of historical catch
GOA	octopus	6	NA	NA
GOA	Sharks - others	6	NA	NA
GOA	squids	6	NA	NA

Table 3. Subset of results showing the mean relative biomass error for combined areas "pollock" like simulations comparing weighted average methods based on 100 simulations for each row. Other simulation cases with variable survey frequencies revealed a similar pattern.

Mean Relative Biomass Error

Factors				Method			
F trend	σ_R	σ_{survey}	Survey frequency	Generalized ARIMA model	Random Effects Model (RE)	Exponential weighting (EW)	Percentage White noise
up/down	0.8	0.15	1	0.334	0.234	0.273	0
up/down	0.4	0.15	1	0.222	0.205	0.231	0
up/down	0.8	0.35	1	0.051	0.104	0.060	0
up/down	0.4	0.35	1	0.016	0.061	0.023	0
down/up	0.8	0.15	1	-0.023	-0.030	-0.009	0
down/up	0.4	0.15	1	0.007	-0.031	-0.009	0
down/up	0.8	0.35	1	-0.025	-0.066	-0.005	1
down/up	0.4	0.35	1	-0.010	-0.065	-0.004	0
flat	0.8	0.15	1	0.033	0.012	0.027	16
flat	0.4	0.15	1	0.024	0.018	0.016	52
flat	0.8	0.35	1	0.058	-0.021	0.028	57
flat	0.4	0.35	1	0.040	0.039	0.056	66

Table 4. Subset of results showing the standard deviation of relative biomass error for combined areas "pollock" like simulations comparing weighted average methods based on 100 simulations for each row. Other simulation cases with variable survey frequencies revealed a similar pattern.

Standard Deviation Relative Biomass error

Factors				Method		
F trend	σ_R	σ_{survey}	Survey frequency	Generalized ARIMA model	Random Effects Model (RE)	Exponential weighting (EW)
up/down	0.8	0.15	1	0.388	0.295	0.349
up/down	0.4	0.15	1	0.292	0.200	0.272
up/down	0.8	0.35	1	0.111	0.094	0.132
up/down	0.4	0.35	1	0.166	0.126	0.152
down/up	0.8	0.15	1	0.133	0.117	0.136
down/up	0.4	0.15	1	0.082	0.104	0.111
down/up	0.8	0.35	1	0.221	0.207	0.231
down/up	0.4	0.35	1	0.195	0.194	0.190
flat	0.8	0.15	1	0.131	0.136	0.137
flat	0.4	0.15	1	0.114	0.105	0.116
flat	0.8	0.35	1	0.210	0.214	0.210
flat	0.4	0.35	1	0.175	0.173	0.247

Table 5. Subset of results showing the mean relative biomass error for combined areas "rockfish" like simulations comparing weighted average methods based on 100 simulations for each row. Other simulation cases with variable survey frequencies revealed a similar pattern.

Mean Relative Biomass Error

Factors				Method			Percentage
F trend	σ_R	σ_{survey}	Survey frequency	Generalized ARIMA model	Random Effects Model (RE)	Exponential weighting (EW)	White noise
up/down	0.8	0.15	1	0.024	0.056	0.037	0
up/down	0.4	0.15	1	0.030	0.070	0.051	0
up/down	0.8	0.35	1	0.103	0.085	0.105	0
up/down	0.4	0.35	1	0.137	0.155	0.157	0
down/up	0.8	0.15	1	-0.032	-0.036	-0.032	0
down/up	0.4	0.15	1	-0.023	-0.031	-0.030	0
down/up	0.8	0.35	1	-0.083	-0.082	-0.073	1
down/up	0.4	0.35	1	-0.047	-0.070	-0.046	0
flat	0.8	0.15	1	0.005	0.008	0.006	16
flat	0.4	0.15	1	0.032	0.023	0.004	52
flat	0.8	0.35	1	-0.008	0.048	0.051	57
flat	0.4	0.35	1	0.022	0.030	0.094	66

Table 6. Subset of results showing the standard deviation of relative biomass error for combined areas "rockfish" like simulations comparing weighted average methods based on 100 simulations for each row. Other simulation cases with variable survey frequencies revealed a similar pattern.

Standard Deviation Relative Biomass error

Factors				Method		
F trend	σ_R	σ_{survey}	Survey frequency	Generalized ARIMA model	Random Effects Model (RE)	Exponential weighting (EW)
up/down	0.8	0.15	1	0.113	0.093	0.106
up/down	0.4	0.15	1	0.121	0.083	0.100
up/down	0.8	0.35	1	0.176	0.182	0.176
up/down	0.4	0.35	1	0.189	0.204	0.170
down/up	0.8	0.15	1	0.091	0.084	0.089
down/up	0.4	0.15	1	0.102	0.096	0.091
down/up	0.8	0.35	1	0.139	0.168	0.126
down/up	0.4	0.35	1	0.156	0.166	0.146
flat	0.8	0.15	1	0.083	0.081	0.084
flat	0.4	0.15	1	0.088	0.066	0.097
flat	0.8	0.35	1	0.160	0.176	0.204
flat	0.4	0.35	1	0.247	0.143	0.359

Figures

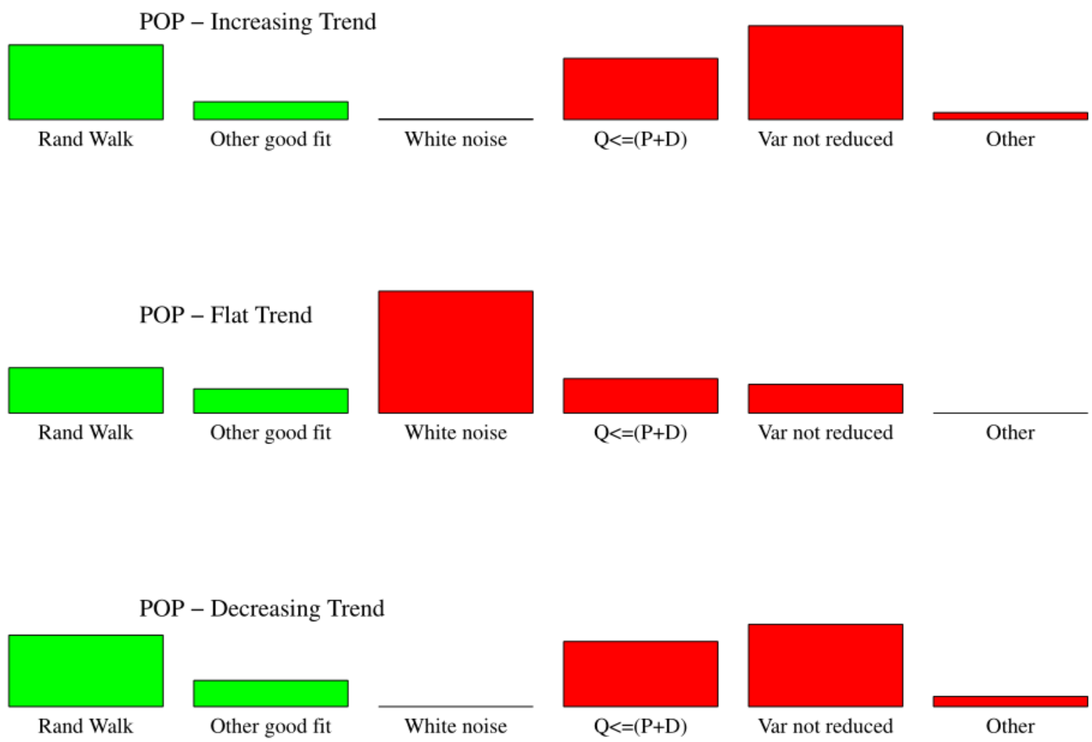
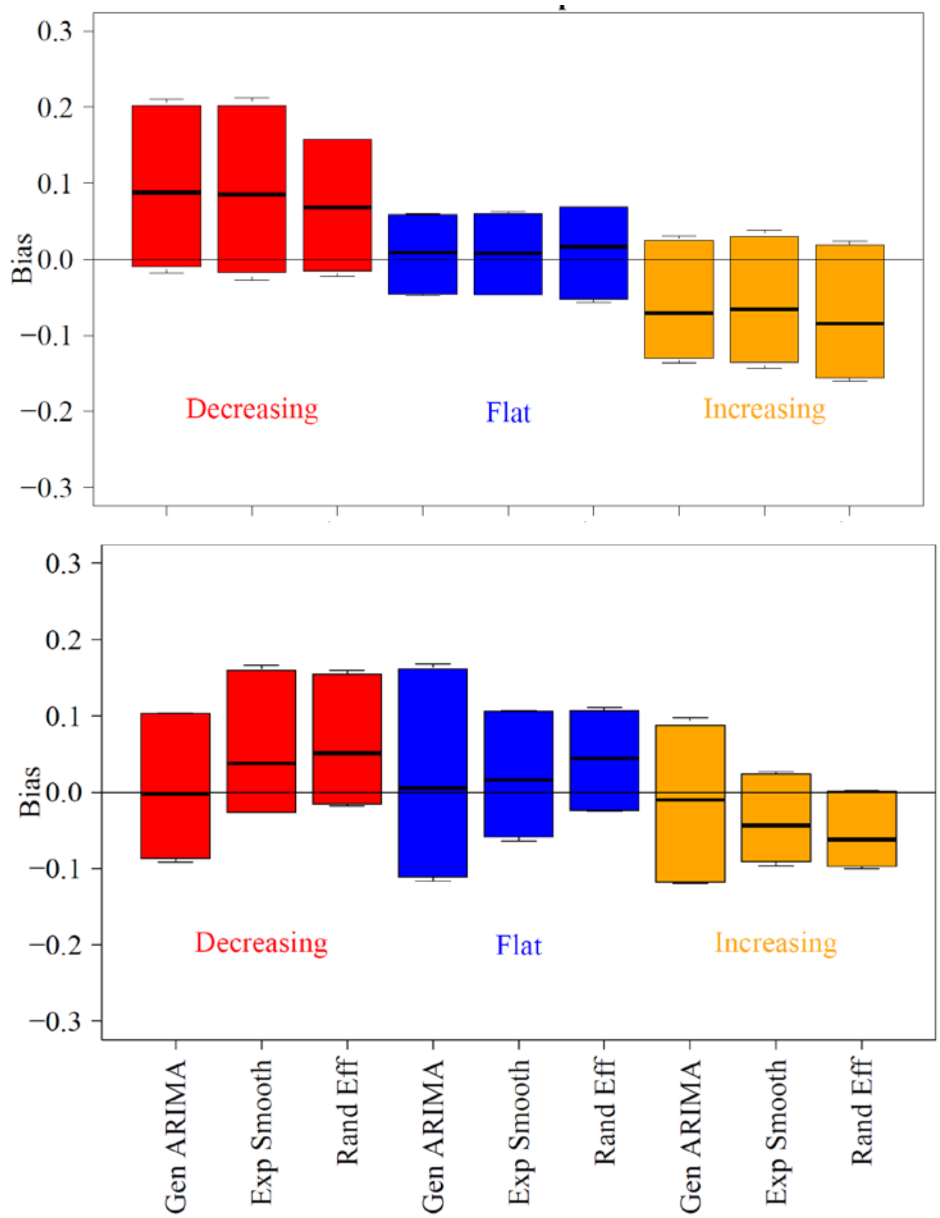


Figure 1. Results of the generalized ARIMA model fitting showing rejected cases (in red) and acceptable models for smoothing for the rockfish-like (POP) simulations with surveys occurring in every year. The panels represent different underlying trends specified in the simulations.



a)

b)

Figure 2. Results comparing bias when the best ARIMA model is (0,1,1) (i.e., a random walk, top panel a) and when the ARIMA fitting procedure indicated that the underlying model was different from a (0,1,1) process (lower panel, b).

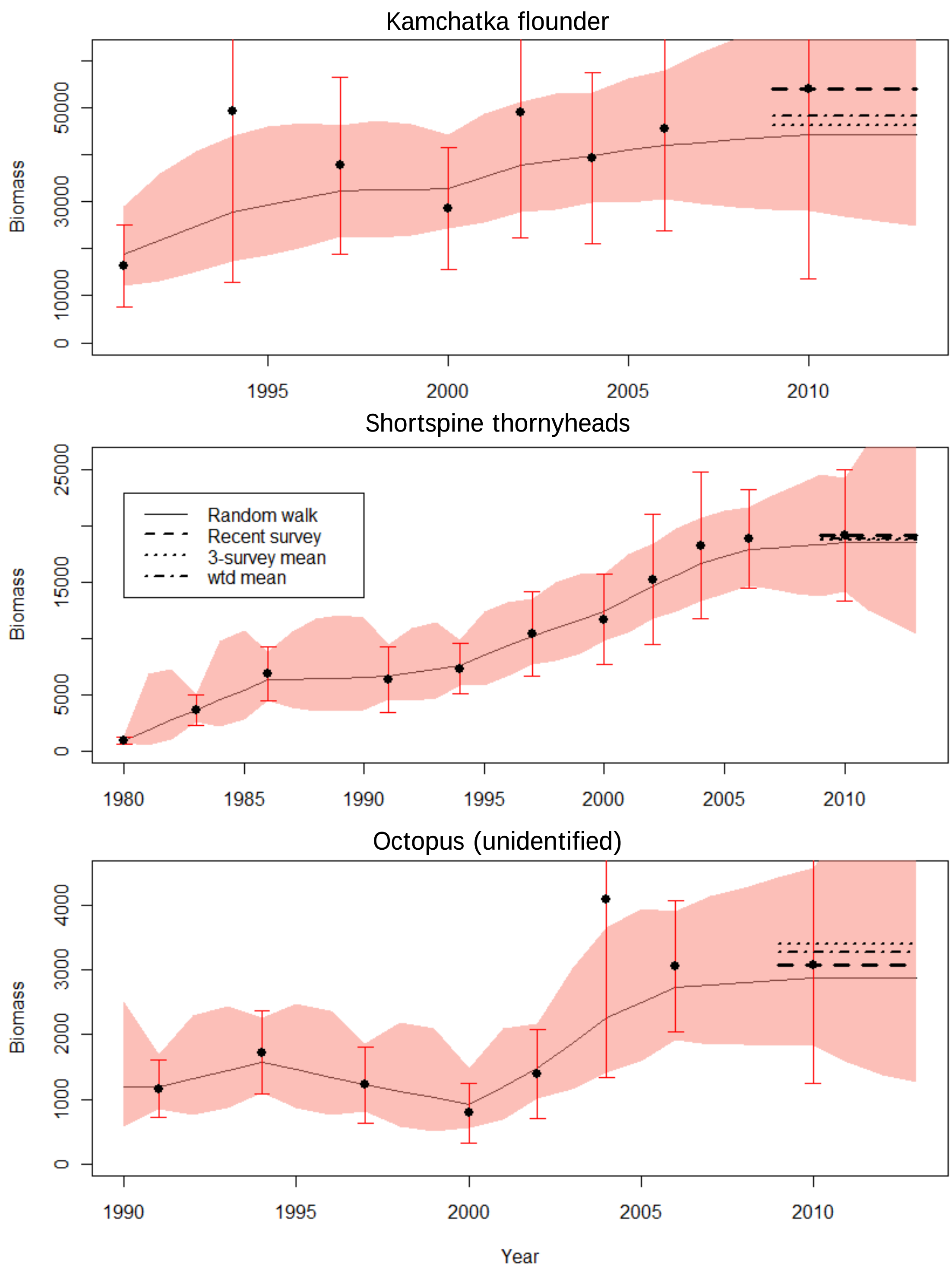


Figure 3. Aleutian Islands survey biomass fits for the random-walk model for some selected stocks. Shaded region represents $\pm$ 2 standard deviations from biomass estimates and error bars on points represents survey (observation) errors.

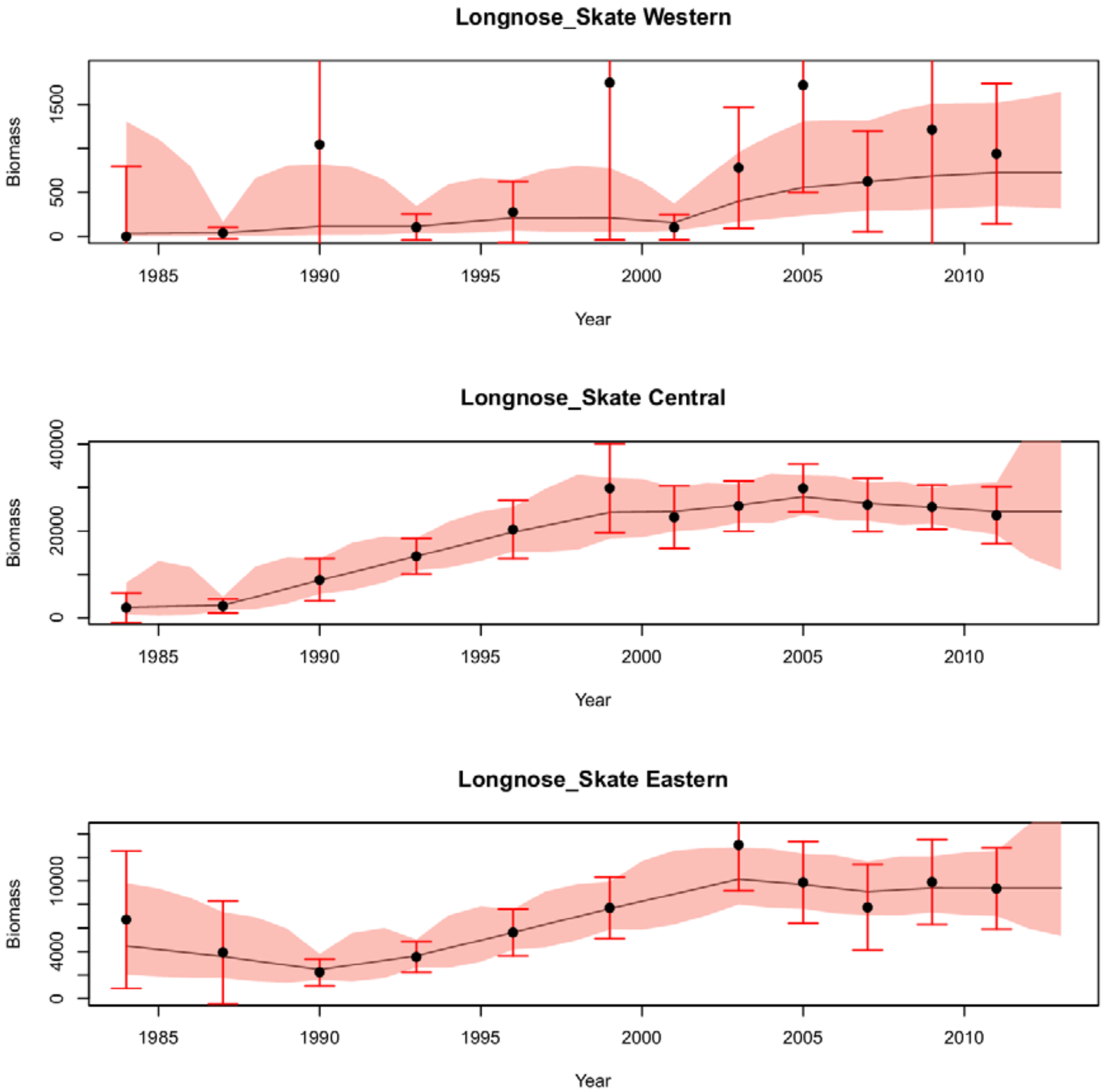


Figure 4. Gulf of Alaska survey biomass fits for the random-walk model for longnose skate showing how 2001 is missing in the eastern region (bottom panel). Shaded region represents $\pm$ 2 standard deviations from biomass estimates and error bars on points represents survey (observation) errors. Note that the vertical scales differ between regions.

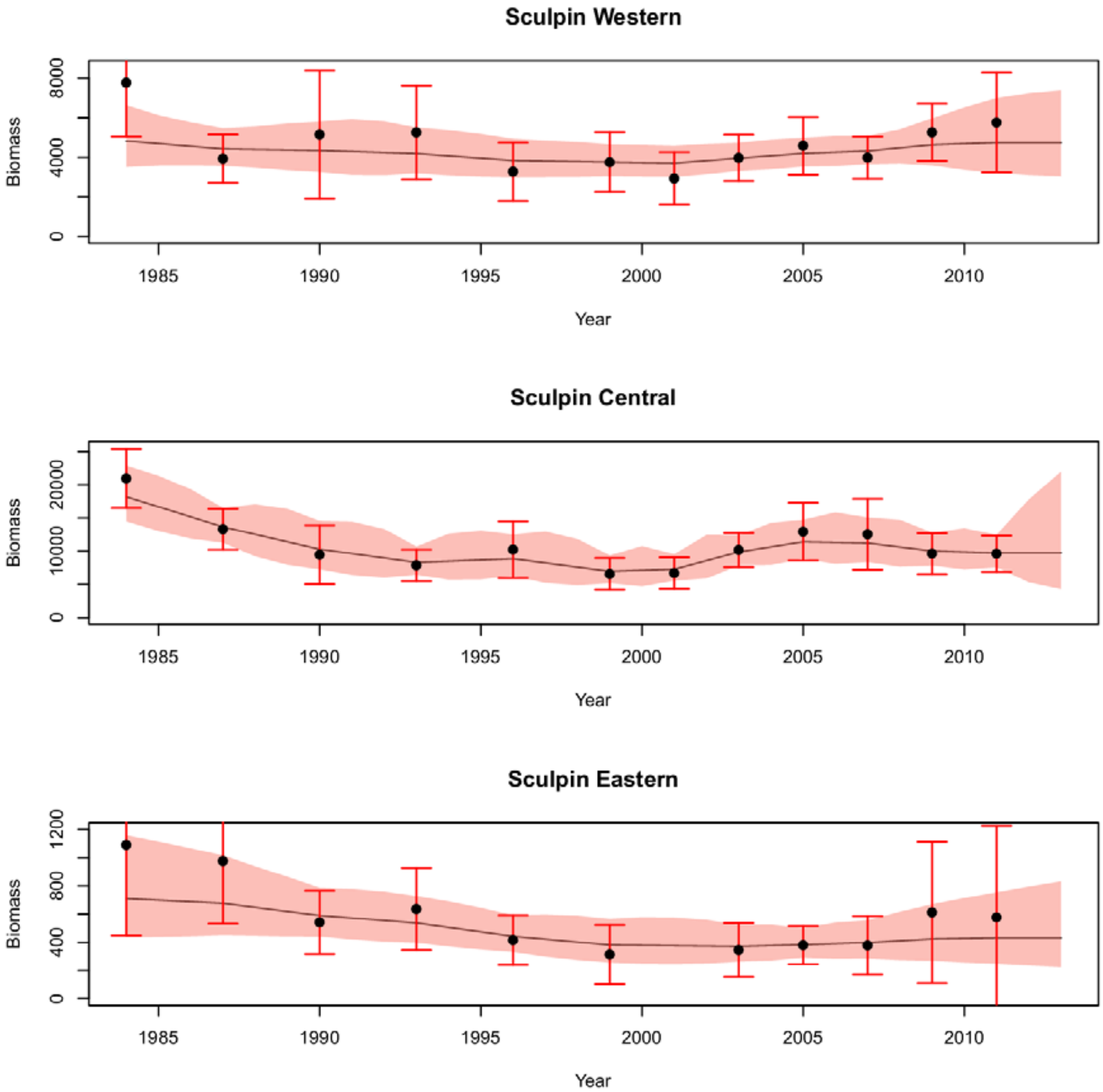


Figure 5. Gulf of Alaska survey biomass fits for the random-walk model for the sculpin complex showing how 2001 is missing in the eastern region (bottom panel). Shaded region represents $\pm$ 2 standard deviations from biomass estimates and error bars on points represents survey (observation) errors. Note that the vertical scales differ between regions.